

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平4-225471

(43) 公開日 平成4年(1992)8月14日

(51) Int.Cl. ⁵	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 15/40	5 1 0 L	7056-5L		
	5 0 0 T	7056-5L		

審査請求 未請求 請求項の数3(全 4 頁)

(21) 出願番号 特願平2-407098

(22) 出願日 平成2年(1990)12月27日

(71) 出願人 000005234

富士電機株式会社

神奈川県川崎市川崎区田辺新田1番1号

(71) 出願人 000237156

富士ファコム制御株式会社

東京都日野市富士町1番地

(72) 発明者 伊東 英俊

東京都日野市富士町1番地 富士ファコム

制御株式会社内

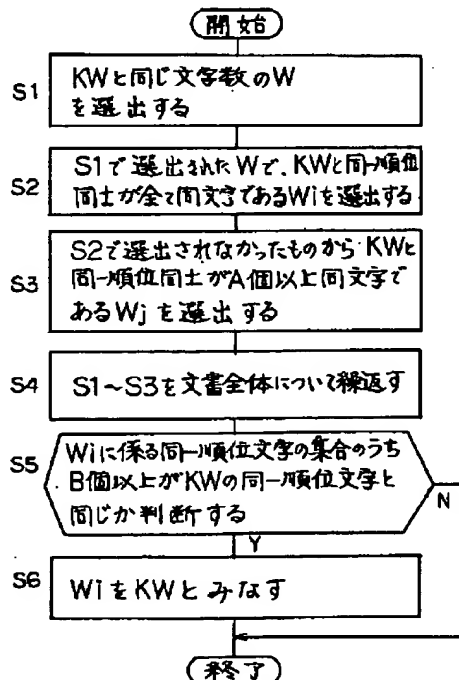
(74) 代理人 弁理士 山口 巖

(54) 【発明の名称】 キーワード検索方法

(57) 【要約】

【目的】 キーワード構成文字の一部に誤読があると、その分だけ検索効率が低下するが、誤読を単純に排除しないで合理的に救済することによって、検索効率の向上を図りたい。

【構成】 読取文字列と登録キーワードとの同一順位の文字同士が登録キーワード文字数に応じた第1所定数以上一致するものを選出し、この選出した各読取文字列について各順位ごとの文字集合を求める。その各順位ごとの文字集合のうちに登録キーワードの対応する順位の文字と一致する文字が、各文字集合に属する文字数に応じた第2所定数以上あるとき、先に選出した読取文字列の全てを検索すべきキーワードとみなす。



1

2

【特許請求の範囲】

【請求項1】登録キーワードと同じ文字列をキーワードとして文書中から検索する方法において、この文書の各文字を文字読取装置によって読み取り、前記登録キーワードと同一文字数で、かつ各同一順位の文字同士が第1の所定数以上一致し、前記登録キーワードの各文字全てとは一致しない読取文字列を選出し；この選出された各読取文字列について各同一順位ごとの文字の集合を求め；この各順位ごとの文字集合のうちに前記登録キーワードの対応する順位の文字と一致する文字の個数が第2

10

の所定数以上あるとき、前記各読取文字列の全てを検索すべきキーワードとみなす；ことを特徴とするキーワード検索方法。

【請求項2】請求項1に記載の方法において、第1所定数は、登録キーワードの文字数に応じて定められることを特徴とするキーワード検索方法。

【請求項3】請求項1または2に記載の方法において、第2所定数は、各順位ごとの文字集合に属する共通な文字数に応じて定められることを特徴とするキーワード検索方法。

20

【発明の詳細な説明】

【0001】

【産業上の利用分野】この発明は、登録キーワードと同じ文字列をキーワードとして文書中から検索する方法であって、とくにキーワード構成文字の一部に誤読があったとしても、単純に排除しないで合理的に救済することにより、検索効率の向上が図れるキーワード検索方法に関する。

【0002】

【従来の技術】一般に、文書の内容を迅速、的確に把握するには、キーワードを活用するのが有効である。たとえば、地球環境保護の問題に関する文書では、たとえば「放射能」や「オゾン層」、「地球汚染」などのキーワードが用いられる。さて、文書のデータベース化をおこなうとき、文書の文字つまり原稿文字を順に文字読取装置によって、標準文字に対する類似度のもっとも高い文字を読取文字として選出し、文字コードで表されるテキストを作成する。このテキストに対して、登録されたキーワードと同じ単一文字または文字列をキーワードとして検索する。

30

【0003】

【発明が解決しようとする課題】従来の方法では、文字読取装置によって読み取られた結果に誤り、つまり誤読が1字でもあると、キーワードが存在するにもかかわらず、検索対象から除外される。すなわち、文書中のキーワード総数に対する検索キーワード数の比率を検索効率と定義したとき、検索効率は著しく低下する。

【0004】この発明の課題は、従来の技術がもつ以上の問題点を解消し、キーワード構成文字の一部に誤読があ

50

ことにより、検索効率の向上が図れるキーワード検索方法を提供することにある。

【0005】

【課題を解決するための手段】この課題を解決するために、請求項1に係るキーワード検索方法は、登録キーワードと同じ文字列をキーワードとして文書中から検索する方法において、この文書の各文字を文字読取装置によって読み取り、前記登録キーワードと同一文字数で、かつ各同一順位の文字同士が第1の所定数以上一致し、前記登録キーワードの各文字全てとは一致しない読取文字列を選出し；この選出された各読取文字列について各同一順位ごとの文字の集合を求め；この各順位ごとの文字集合のうちに前記登録キーワードの対応する順位の文字と一致する文字の個数が第2の所定数以上あるとき、前記各読取文字列の全てを検索すべきキーワードとみなす。

【0006】請求項2に係るキーワード検索方法は、請求項1に記載の方法において、第1所定数は、登録キーワードの文字数に応じて定められる。

【0007】請求項3に係るキーワード検索方法は、請求項1または2に記載の方法において、第2所定数は、各順位ごとの文字集合に属する共通な文字数に応じて定められる。

【0008】

【作用】請求項1に係るキーワード検索方法では、①文書の各文字を文字読取装置によって読み取り、登録キーワードと同一文字数で、かつ各同一順位の文字同士が第1の所定数以上一致し、登録キーワードの各文字全てとは一致しない読取文字列を選出する、つまり登録キーワードと部分的に一致し、従来は除外されるべき読取文字列について救済可能な候補として1次選考する、②この選出された各読取文字列について各同一順位ごとの文字の集合を求め、この各順位ごとの文字集合のうち登録キーワードの対応する順位の文字と一致する文字の個数が第2の所定数以上あるとき、2次選考として各読取文字列の全てを検索すべきキーワードとみなす。なお、1次選考における第1所定数は、請求項2のように登録キーワードの文字数に応じて、また2次選考における第2所定数は、請求項3のように各順位ごとの文字集合に属する共通な文字数に応じてそれぞれ定められる。

40

【0009】

【実施例】本発明に係るキーワード検索方法を適用した検索装置について、以下に図を参照しながら説明する。図3は検索装置に係る登録キーワードと1次選考読取文字列の例示図である。図3において、登録キーワードKは、4文字から構成される「富士電機」である。1次選考の結果、5個の読取文字列W1～W5が選出されたとする。すなわち、富士電機、富士電機、富士電機、富士電機、富士電機 である。各読取文字列とも登録キーワードと、下線を付けた2個の同一文字をもっている。こ

こで、発明における第1所定数は2とする。

【0010】図4は読取文字列の2次選考に係る選考過程図である。図4において、第1列に文字順位、第2列に登録キーワード、第3列に1次選考読取文字列の各順位文字の集合、第4列に登録キーワードの各構成文字と各順位文字の集合との一致文字数、がそれぞれ示される。たとえば、文字順位1に相当する登録キーワードの構成文字は「富」、これに対して1次選考された5個の読取文字列W1～W5で文字順位1に相当する文字の集合は{富、富、笛、宮、宙}である。つまり、一致文字数は2である。ここで、発明における第2所定数は2とする。同様に、各文字順位2, 3, 4について文字集合を求め、各一致文字数2, 2, 4を得る。したがって、1次選考された5個の読取文字列W1～W5は、2次選考にも合格してキーワードであると判定される。なお、第1, 第2の各所定数は、基本的には経験的に定められ、一般的には、第1所定数は登録キーワードの構成文字数が多くなるほど大きい数値をとり、第2所定数は各順位の文字集合に属する文字数が多くなるほど比例的に大きい数値をとる。

【0011】図2は検索装置の構成を示すブロック図である。図2において、1は文書の原稿文字に係る画像を求めるイメージスキャナ、2は読取部で、原稿文字に係る画像に基づいて読み取りをおこなう。3は読取文字用のメモリで、ここに読取文字が文字コードで格納される。4は登録キーワード用の入力部、5は登録キーワード用のメモリである。6は照合部で、各メモリ3, 5からの対応する文字コードを照合し、一致、不一致の判定をする等、前記の1次、2次の各選考処理をおこない、最終的にキーワードとみなすかどうかを決める。7はC

R Tで、照合結果を画面に表示する。なお、このC R T 7に照合結果を印刷して出力するプリンタを併設することができる。

【0012】図1は検索装置の動作を示すフローチャートである。図1のステップS1で、キーワードKWと同じ文字数の読取文字列Wを順次選出(予備選考)する。ステップS2で、S1で選出された読取文字列Wで、KWと同一順位同士が全て同文字である読取文字列Wiを選出する。これは従来の検索方法である。ステップS3で、S2で選出されなかった読取文字列から、KWと同一順位同士がA個(第1所定数)以上、同文字である読取文字列Wjを選出(1次選考)する。すなわち、ステップS3以降が検索キーワードの救済処置になる。なお、図3の例では、A=2である。ステップ4で、ステップS1～S3を文書全体について繰り返す。ステップS5で、S3で選出された読取文字列Wjに係る同一順位文字の集合のうちB個(第2所定数)以上が、KW

の同一順位文字と同かどうか判断(2次選考)し、YESならステップS5に移行し、NOなら救済されず終了する。なお図4の例では、B=2である。ステップS6で、読取文字列WjをKWとみなし、救済する。したがって、検索キーワードは各読取文字列Wi, Wjになる。

【0013】以上のように、従来の方法で除外された読取文字列を、1次、2次の各選考過程を経て救済するが、この救済が適正かつ合理的な制約条件のもとでおこなわれるから、救済により検索効率の向上が図れるとともに、救済された読取文字列の検索確度は高い。

【0014】

【発明の効果】請求項1～3のいずれかに係るキーワード検索方法では共通に、登録キーワードと同一文字数で、かつ各同一順位の文字同士が第1の所定数以上一致し、登録キーワードの各文字全てとは一致しない読取文字列を1次選考として選出し；この選出された各読取文字列について各同一順位ごとの文字の集合を求め、この各順位ごとの文字集合のうち登録キーワードの対応する順位の文字と一致する文字の個数が第2の所定数以上あるとき、2次選考として各読取文字列の全てを検索すべきキーワードとみなす。

【0015】したがって、キーワード構成文字の一部に誤読があったとしても、単純に排除しないで合理的に救済することにより、検索効率の向上が図れる、という効果が得られる。また、とくに請求項2のように、第1所定数が登録キーワード文字数に応じて定められ、また請求項3のように、第2所定数が各順位ごとの文字集合に属する共通な文字数に応じて定められるから、救済が適正かつ合理的な制約条件のもとでおこなわれ、救済された読取文字列の検索確度は高い。

【図面の簡単な説明】

【図1】本発明に係る方法を適用した検索装置の動作を示すフローチャート

【図2】この検索装置の構成を示すブロック図

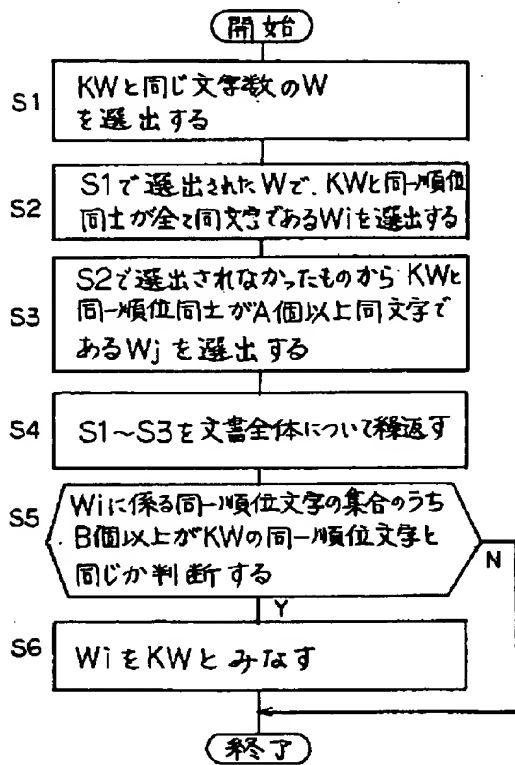
【図3】この検索装置に係る登録キーワードと1次選考読取文字列の例示図

【図4】読取文字列の2次選考に係る選考過程図

【符号の説明】

- 1 イメージセンサ
- 2 読取部
- 3 メモリ
- 4 入力部
- 5 メモリ
- 6 照合部
- 7 C R T

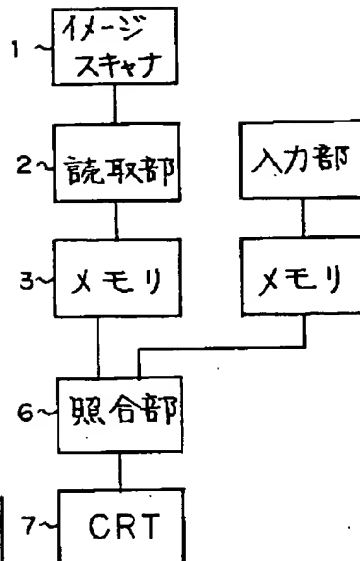
【図1】



【図4】

文字 順位	登録 キーワード	1次選考読取文字列の 各順位文字の集合					一致 文字数
1	富	富	富	富	富	富	2
2	士	士	士	士	士	士	2
3	電	謳	電	雷	壇	電	2
4	機	機	機	機	機	機	4

【図2】



【図3】

K: 富士電機
 W1: 富士謳機
 W2: 富士電機
 W3: 富士雷機
 W4: 富士壇機
 W5: 富士電機